

Editorial pubs.acs.org/jcim

Exploring Novel Directions in Free Energy Calculations



Cite This: J. Chem. Inf. Model. 2020, 60, 5283-5286



ACCESS

III Metrics & More



Article Recommendations

Recognizing the recent developments and applications of free energy methods, the Journal of Chemical Information and Modeling extended an open invitation to the computational chemistry community to contribute to a special issue on Novel Directions in Free Energy Methods and Applications. We are delighted to have received a breadth of manuscripts using free energy calculations, covering diverse methods and applications in alchemical free energy calculations, free energy simulations contributing to kinetics estimation, machine learning to estimate free energy differences and hydration free energies, estimation of conformational free energy, benchmarking studies, and new software applications.

Alchemical free energy calculations typically use a coupling parameter, λ , to split the phase spaces of two end states in intermediate states (windows), which are then sampled by linear mixing the force field parameters or energies of the end states. However, the choice of the λ spacing parameter may affect the entire calculation by impacting energy barriers and the state phase space overlap. König et al.2 provide an alternative method to overcome this problem by using a combination of λ -coupling with the enveloping distribution sampling (EDS) method. They demonstrate the methodology on five benchmark systems and for both relative and absolute free energy calculations, showing that λ -EDS allows for larger steps along the alchemical pathway than conventional intermediate λ states. To avoid the problem of introducing the λ coupling parameter and artificial intermediate states, Perthold et al.³ enhance one-step perturbation methods by using accelerated EDS (A-EDS) by proposing an improved algorithm for the A-EDS Hamiltonian parameter search, which further simplifies the construction of the reference-state Hamiltonian. The authors demonstrate the automated calculation of multiple free-energy differences between different ligands from a single simulation and apply it to three protein-ligand systems including systems with net charge change, previously considered too challenging for multistate EDS approaches. Another consideration for alchemical free energy calculations is that their accuracy depends on a reliable starting structure being available through biophysical methods such as X-ray crystallography or NMR. To tackle this issue, Oshima et al.4 introduce a sequential prediction protocol using generalized replica exchange with solute tempering (gREST) and free energy perturbation (FEP) calculations. They find that this approach helps to accelerate ligand reorientation or rotation in the binding site and also makes the protein binding site residues more flexible and facilitates faster water diffusion near the binding pocket, requiring only a small number of replicas for convergence of the binding pose. This protocol

may aid in predicting binding affinities of ligands without highresolution structural information of the bound state.

The estimation of absolute binding free energies in protein ligand systems is central in computer-aided drug design, yet it poses many challenges because of the amount of sampling needed to annihilate the ligands. Gilabert et al.⁵ present a multistep protocol combining Monte Carlo sampling to identify relevant ligand states of the unbinding process and extensive molecular dynamics (MD) simulations followed by Markov state modeling, which is subsequently used to estimate the binding free energy. Overall, the protocol is capable of properly ranking the set of ligands examined, albeit with a significant computational cost. To rationally design inhibitors with improved steric contacts and enhanced binding free energies, Wade and Huggins⁶ present a new method that uses alchemical single step perturbation calculations to rapidly optimize the van der Waals interactions of a small molecule in a protein-ligand complex in order to maximize its binding affinity. This optimization identifies favorable growth vectors in the ligand, where ligand growth such as methyl scans can be performed with a more than 10-fold speedup compared to traditional free energy calculations and sublinear scaling with the number of growth vectors assessed. To estimate the free energy of predicted protein-protein complexes, Zacharias et al. extended a replica-exchange (RE) based scheme employing different levels of repulsive biasing between partners in each replica simulation (repulsive scaling (RS)-REMD) to simultaneously refine and score protein-protein complexes. The new RS-REMD implementation allows for performing simulations in explicit solvent and, in addition, also allows for extraction of the binding free energy. A new multiscale method called multiscale cell correlation (MCC) is presented to calculate the entropic component of the free energy of proteins from MD simulations.8 The method decomposes the protein into sets of rigid-body units based on their covalent bond connectivity at three levels of hierarchy-molecule, residue, and united atoms—and evaluates the vibrational and topographical entropy from forces, torques, and dihedrals at each level, taking into account correlations between sets of constituent units that together make up a larger unit at the coarser length scale. MCC estimates entropies in close agreement with

Special Issue: Novel Directions in Free Energy Methods and Applications

Published: November 23, 2020





normal-mode analysis and \sim 40% smaller than those calculated using quasiharmonic analysis. Furthermore, it provides much faster convergence.

Approximate free energy methods for the estimation of protein-ligand affinities such as MM/GBSA have gained popularity due to their limited computational cost; however, these suffer from accuracy limitations of the standard generalized Born (GB) model. In their study, Wang et al.9 present and evaluate an MM/GBSA approach based on a variable dielectric generalized Born (VDGB) model using residue-type-based dielectric constants for predicting proteinligand binding affinities. In the VDGB model, different dielectric values were assigned to the three types of protein residues and a new charge-dependent dielectric model was proposed for small ligands. Free energy methods are also widely used in mapping the free energy landscape of molecular processes as well as finding an optimal pathway connecting minima on a multidimensional free-energy landscape. To this end, Fu et al. 10 combine a well-tempered metadynamicsextended adaptive biasing force (WTM-eABF) scheme with a path-searching algorithm (MULE), to identify the underlying minimum free-energy pathway in the collective-variable space of interest. The binding free energy is not the only crucial parameter that defines ligand activity. In fact, ligand binding kinetics often correlate with the ligand efficacy in vivo and are an essential parameter in lead optimization. Ahn et al. 11 used a weighted ensemble approach called the concurrent adaptive sampling (CAS) algorithm that uses MD simulations to rank seven ligands for β -cyclodextrin. They compared the performance of CAS to another multiscale milestoning approach called simulation enabled estimation of kinetic rates (SEEKR), which uses both MD and Brownian dynamics simulations. Similarities and differences of the SEEKR and CAS algorithms and when to use each method for the receptor-ligand system of interest are discussed.

The relevance and impact of free-energy methods was showcased on a range of challenging systems. Guest et al. 12 used MD simulations and FEP calculations to investigate potential antagonists of the arginine-glycine-aspartic acid (RGD) binding integrin $\alpha v \beta 6$, which is linked to the initiation and progression of chronic lung disease idiopathic pulmonary fibrosis. Using MD/FEP, the importance of hydrogen bonding and cation- π interactions for binding were highlighted. Deflorian et al. 13 developed a protocol to perform accurate FEP calculations of ligand binding affinities for GPCRs. GPCRs represent one of the most important drug target classes but are notoriously challenging to study with computational methods. In this study, the authors established the important role of key water molecules, amino acid ionization states and careful equilibration protocols for successful free energy calculations. Another important system is the hemoglobinase falcipain-2, which is a potential drug target for combating malaria. With MD simulations and free-energy calculations, Hernández-González et al. 14 elucidated the structural determinants for the high affinity and selectivity of peptidomimetic nitriles. They found that water bridge formation or displacement of water bridges need to be taken into account in the rational design of selective inhibitors. MD-based calculation of relative free energies can also be used to investigate chirality effects of small organic molecules. Plazinska et al. 15 compared four different enhanced sampling strategies to enable the interconversion between stereoisomers. The free energy profiles corresponding to the inversion of the chiral center are recovered by applying coordinate-based methods and can be subsequently used as a time-independent bias allowing one to sample stereogenic interconversion within relatively short simulation times and for the simultaneous treatment of multiple chiral centers.

MD simulations together with approximate free energy methods were used to evaluate a loop grafting approach to achieve "silencing" of the fragment crystallizable (Fc) region of monoclonal antibodies, i.e. a reduction of the affinity of this region with immune receptors. 16 The binding of the Fc variants to the receptor was predicted using MM/GBSA and it was shown that loop grafting demonstrates a high potential as a new approach for the design of stable Fc-silent variants that could be used as building blocks of novel antibody therapeutics. Goossens et al.¹⁷ investigated the structural behavior and reaction mechanism of the glycosyltransferase of Staphylococcus aureus in order to aid targeted design of new antibiotics. The large-scale dynamics of the enzyme-ligand complex were probed with MD simulations, and QM/MM calculations were used to propose a reaction mechanism. The value of MD simulations as a key component to refine docking results was showcased with modulators of the membraneassociated protein kinase C.18 The unexpected poor experimental activity of a docking hit from a previous study could be rationalized using MD simulations and thermodynamic integration (TI) in the relevant environment.

Machine learning (ML) is becoming more and more widespread in chemistry in recent years, and its use is apparent in the many manuscripts submitted for quantum chemistry and MD simulations in this special issue. Bennett et al. 19 performed atomistic MD simulations of 1500 small molecules in a water/ cyclohexane system to obtain the transfer free energy and then expanded the calculations to 15 000 molecules using TI. This set was subsequently used to train a neural network (NN) model (spatial graph NN or 3D-convolutional NN) to predict transfer free energies. Gebhardt et al.²⁰ also used MD simulations as input to train a ML model. In this case, short simulations of small molecules in water and as pure liquid were carried out, and the trajectory information was encoded in socalled MD fingerprints. With these, a gradient tree boosting model was trained to predict self-solvation free energies and limiting activity coefficients. In a third approach, machine learning was used to introduce an empirical correction term for FEP simulations to estimate hydration free energies.²¹ The deviation between the calculated and experimental hydration free energy is used as target property to train the ML models, which can then be employed to correct the calculated FEP results of new compounds.

Free energy calculations have been applied in drug discovery for the last few decades. Song et al. 22 presents a detailed overview of the evolution of these types of calculations and how they have impacted the drug discovery process. Alchemical free energy calculations have been made possible by the development of force fields, sampling techniques, methods, and hardware and are continuously being updated. New and updated methods have been implemented in commercial software packages, which have been described in this special issue. The accuracy of FEP relies on protein preparation and system setup, as described by Shih et al. 23 NF-kappa-B-inducing kinase (NIK) was examined to assess the accuracy of Schrödinger's FEP+ platform for free energy calculations. Three independent computational chemists (with and without kinase experience) performed protein preparation

on the NIK protein and FEP was run on five ligands with a dynamic range in activity. Only one system setup accurately reproduced all five actives. Careful inspection of the protein and water network was found to be necessary for accurate predictions using FEP. The Merck KGaA group also highlighted the importance of many different factors that influence the accuracy of the free energy calculations in a typical drug discovery context by looking at both internal targets and a large benchmarking set. For the targets where the desired accuracy was not achieved, results were still deemed useful to prioritize compounds for synthesis. The team also found that FEP calculations were most relevant in the hitto-lead or fragment optimization stages of drug discovery to identify compounds with more optimized potency.

The work by these two teams^{23,24} was performed in an industrial setting, using Schrödinger's FEP+ methodology. There are also many academic software packages that have become more widely used and are being integrated into commercial drug discovery platforms. One of those is AMBER's implementation of TI, which has been highlighted by two manuscripts in this special issue. First is an article by Tsai et al.²⁵ that describes validation of free energy calculations and methods in a modified version of AMBER18 (now AMBER20) using a GPU-accelerated code. A unified and split protocol was simulated to understand inconsistencies seen in the TI implementation in AMBER16. While both protocols yielded consistent results, the authors demonstrated the need for proper treatment of the electrostatic and van der Waals interactions between the soft and common core regions of the molecule. AMBER's TI implementation has not been adopted in traditional industrial drug discovery efforts as quickly as other methods due to inconsistencies in the treatment of soft core potentials,²⁶ for workflows for alchemical transformations and for postprocessing analysis. An industry-academia collaboration was formed to address some of these known issues and to highlight best practices for running free energy calculations (protein preparation, treatment of water molecules, ligand preparation, atom mapping, λ schedule, etc.). Additionally, the authors suggested that a continuing industry-academia collaboration will be necessary to implement novel free energy developments within AMBER. Chen et al.27 described a hardware update for NAMD that allows for GPU-accelerated FEP calculations providing speedups of 30fold over the CPU implementation. Four test systems were used to confirm reliability, accuracy and performance of the GPU-based code, and the GPU-based simulations experienced 4-30-fold speedups depending on the GPU used.

Method developments in software for free energy calculations were also a theme in this special issue. The movable type (MT) method, previously developed by the Merz group, has been implemented into Quantum Bio Inc.'s DivCon Discovery Suite.²⁸ MT calculates the atomic partition function from an input geometry of a ligand and then uses this to estimate the free energy of that ligand. End-state (single pose) and ensemble (multiple docked poses) scoring was used to assess the accuracy and speed of the method for the CASF-2016 (comparative assessment of scoring functions) set (57 proteins, 285 ligands), and the HPF (homology protein family) set (10 proteins, 248 ligands). For most of these validation sets, the mean unsigned error was reported to be below 1 kcal/mol, and more importantly, each calculation took approximately 25 min/ligand on a single CPU, compared to traditional FEP calculations that take a few hours on a single

GPU. Another platform illustrated in this special issue is the biomolecular reaction and interaction dynamics global environment (BRIDGE), which is an open-source web-based platform that has implemented FEP-based methods from GROMACS and YANK to calculate both absolute and relative binding free energies (ABFE, RBFE, respectively).²⁹ The interaction module in BRIDGE allows for automation of system setup, simulation and postprocessing, and is easily transferable and accessible in GitHub. The authors demonstrated this transferability by independently calculating the free energies of multiple ligands bound to CDK2 in two research groups and demonstrated reproducibility of the independent results (within 1 kcal/mol). A second test case was described to demonstrate reproducibility for ligands bound to ST3GaI-I. Here, the authors used ABFE simulations to accurately predict the free energy of the ligands. ABFE calculations are often difficult to setup and run, and with the BRIDGE platform, these simulations can be easily reproduced by many groups across the world.

2020 has been an extraordinary year so far. We were very encouraged by the number of high-quality manuscripts received for the special issue on Novel Directions in Free Energy Methods and Applications amidst the global pandemic crisis. In total, 28 manuscripts comprise the special issue and exemplify the many ways free energy simulations are used today. The special issue also highlights the prospective applications and further development of free energy methods as research groups integrate machine learning and other methodologies into the free energy field. We are pleased to showcase the considerable efforts performed by many research groups in academia and industry and are excited about what the future holds in free energy methodology and applications.

Kira A. Armacost © orcid.org/0000-0001-9322-7981 Sereina Riniker © orcid.org/0000-0003-1893-4031 Zoe Cournia © orcid.org/0000-0001-9287-364X

AUTHOR INFORMATION

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.0c01266

Notes

Views expressed in this editorial are those of the authors and not necessarily the views of the ACS.

The authors declare the following competing financial interest(s): K.A.A. is a former employee of Merck & Co., Inc., Whitehouse Station, NJ, U.S., and potentially owns stock and/or holds stock options in the Company.

ACKNOWLEDGMENTS

The authors would like to thank Juan Alvarez, John Sanders, and Xin Cindy Yan for their review of this editorial.

REFERENCES

- (1) Armacost, K. A.; Riniker, S.; Cournia, Z. Novel Directions in Free Energy Methods and Applications. *J. Chem. Inf. Model.* **2020**, *60* (1), 1–5.
- (2) König, G.; Glaser, N.; Schroeder, B.; Kubincová, A.; Hünenberger, P. H.; Riniker, S. An Alternative to Conventional λ -Intermediate States in Alchemical Free Energy Calculations: λ -Enveloping Distribution Sampling. *J. Chem. Inf. Model.* **2020**, DOI: 10.1021/acs.jcim.0c00520.
- (3) Perthold, J. W.; Petrov, D.; Oostenbrink, C. Toward Automated Free Energy Calculation with Accelerated Enveloping Distribution

- Sampling (A-EDS). J. Chem. Inf. Model. 2020, DOI: 10.1021/acs.jcim.0c00456.
- (4) Oshima, H.; Re, S.; Sugita, Y. Prediction of Protein-Ligand Binding Pose and Affinity Using the gREST+FEP Method. *J. Chem. Inf. Model.* **2020**, DOI: 10.1021/acs.jcim.0c00338.
- (5) Gilabert, J. F.; Gracia Carmona, O.; Hogner, A.; Guallar, V. Combining Monte Carlo and Molecular Dynamics Simulations for Enhanced Binding Free Energy Estimation through Markov State Models. J. Chem. Inf. Model. 2020, DOI: 10.1021/acs.jcim.0c00406.
- (6) Wade, A. D.; Huggins, D. J. Identification of Optimal Ligand Growth Vectors Using an Alchemical Free-Energy Method. *J. Chem. Theory Comput.* **2020**, DOI: 10.1021/acs.jctc.0c00610.
- (7) Siebenmorgen, T.; Zacharias, M. Efficient Refinement and Free Energy Scoring of Predicted Protein-Protein Complexes Using Replica Exchange with Repulsive Scaling. *J. Chem. Inf. Model.* **2020**, DOI: 10.1021/acs.jcim.0c00853.
- (8) Chakravorty, A.; Higham, J.; Henchman, R. H. Entropy of Proteins Using Multiscale Cell Correlation. *J. Chem. Inf. Model.* **2020**, DOI: 10.1021/acs.jcim.0c00611.
- (9) Wang, E.; Liu, H.; Wang, J.; Weng, G.; Sun, H.; Wang, Z.; Kang, Y.; Hou, T. Development and Evaluation of MM/GBSA Based on a Variable Dielectric GB Model for Predicting Protein-Ligand Binding Affinities. J. Chem. Inf. Model. 2020, DOI: 10.1021/acs.jcim.0c00024.
- (10) Fu, H.; Chen, H.; Wang, X. A.; Chai, H.; Shao, X.; Cai, W.; Chipot, C. Finding an Optimal Pathway on a Multidimensional Free-Energy Landscape. *J. Chem. Inf. Model.* **2020**, DOI: 10.1021/acs.jcim.0c00279.
- (11) Ahn, S.-H.; Jagger, B. R.; Amaro, R. E. Ranking of Ligand Binding Kinetics Using a Weighted Ensemble Approach and Comparison with a Multiscale Milestoning Approach. *J. Chem. Inf. Model.* **2020**, DOI: 10.1021/acs.jcim.9b00968.
- (12) Guest, E. E.; Oatley, S. A.; Macdonald, S. J. F.; Hirst, J. D. Molecular Simulation of $\alpha v \beta 6$ Integrin Inhibitors. *J. Chem. Inf. Model.* **2020**, DOI: 10.1021/acs.jcim.0c00254.
- (13) Deflorian, F.; Perez-Benito, L.; Lenselink, E. B.; Congreve, M.; van Vlijmen, H. W. T.; Mason, J. S.; Graaf, C. D.; Tresadern, G. Accurate Prediction of GPCR Ligand Binding Affinity with Free Energy Perturbation. *J. Chem. Inf. Model.* **2020**, DOI: 10.1021/acs.jcim.0c00449.
- (14) Hernández González, J. E.; Hernández Alvarez, L.; Leite, V. B. P.; Pascutti, P. G. Water Bridges Play a Key Role in Affinity and Selectivity for Malarial Protease Falcipain-2. *J. Chem. Inf. Model.* **2020**, DOI: 10.1021/acs.jcim.0c00294.
- (15) Plazinska, A.; Plazinski, W. Chirality Effects in Biomolecular Systems: Calculation of the Relative Free Energies by Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2020**, DOI: 10.1021/acs.jcim.0c00605.
- (16) Lešnik, S.; Hodošček, M.; Podobnik, B.; Konc, J. Loop Grafting between Similar Local Environments for Fc-Silent Antibodies. *J. Chem. Inf. Model.* **2020**, DOI: 10.1021/acs.jcim.9b01198.
- (17) Goossens, K.; Neves, R. P. P.; Fernandes, P. A.; De Winter, H. A Computational and Modeling Study of the Reaction Mechanism of Staphylococcus aureus Monoglycosyltransferase Reveals New Insights on the GT51 Family of Enzymes. *J. Chem. Inf. Model.* **2020**, DOI: 10.1021/acs.jcim.0c00377.
- (18) Lautala, S.; Provenzani, R.; Koivuniemi, A.; Kulig, W.; Talman, V.; Róg, T.; Tuominen, R. K.; Yli-Kauhaluoma, J.; Bunker, A. Rigorous Computational Study Reveals What Docking Overlooks: Double Trouble from Membrane Association in Protein Kinase C Modulators. J. Chem. Inf. Model. 2020, DOI: 10.1021/acs.jcim.0c00624.
- (19) Bennett, W. F. D.; He, S.; Bilodeau, C. L.; Jones, D.; Sun, D.; Kim, H.; Allen, J. E.; Lightstone, F. C.; Ingólfsson, H. I. Predicting Small Molecule Transfer Free Energies by Combining Molecular Dynamics Simulations and Deep Learning. *J. Chem. Inf. Model.* **2020**, DOI: 10.1021/acs.jcim.0c00318.
- (20) Gebhardt, J.; Kiesel, M.; Riniker, S.; Hansen, N. Combining Molecular Dynamics and Machine Learning to Predict Self-Solvation

- Free Energies and Limiting Activity Coefficients. *J. Chem. Inf. Model.* **2020**, DOI: 10.1021/acs.jcim.0c00479.
- (21) Scheen, J.; Wu, W.; Mey, A. S. J. S.; Tosco, P.; Mackey, M.; Michel, J. Hybrid Alchemical Free Energy/Machine-Learning Methodology for the Computation of Hydration Free Energies. *J. Chem. Inf. Model.* **2020**, DOI: 10.1021/acs.jcim.0c00600.
- (22) Song, L. F.; Merz, K. M. Evolution of Alchemical Free Energy Methods in Drug Discovery. *J. Chem. Inf. Model.* **2020**, DOI: 10.1021/acs.jcim.0c00547.
- (23) Shih, A. Y., Hack, M.; Mirzadegan, T. Impact of Protein Preparation on Resulting Accuracy of FEP Calculations. *J. Chem. Inf. Model.* **2020**, DOI: 10.1021/acs.jcim.0c00445.
- (24) Schindler, C. E. M.; Baumann, H.; Blum, A.; Böse, D.; Buchstaller, H.-P.; Burgdorf, L.; Cappel, D.; Chekler, E.; Czodrowski, P.; Dorsch, D.; Eguida, M. K. I.; Follows, B.; Fuchß, T.; Grädler, U.; Gunera, J.; Johnson, T.; Jorand Lebrun, C.; Karra, S.; Klein, M.; Knehans, T.; Koetzner, L.; Krier, M.; Leiendecker, M.; Leuthner, B.; Li, L.; Mochalkin, I.; Musil, D.; Neagu, C.; Rippmann, F.; Schiemann, K.; Schulz, R.; Steinbrecher, T.; Tanzer, E.-M.; Unzue Lopez, A.; Viacava Follis, A.; Wegener, A.; Kuhn, D. Large-Scale Assessment of Binding Free Energy Calculations in Active Drug Discovery Projects. *J. Chem. Inf. Model.* 2020, DOI: 10.1021/acs.jcim.0c00900.
- (25) Tsai, H.-C.; Tao, Y.; Lee, T.-S.; Merz, K. M.; York, D. M. Validation of Free Energy Methods in AMBER. *J. Chem. Inf. Model.* **2020**, DOI: 10.1021/acs.jcim.0c00285.
- (26) Lee, T.-S.; Allen, B. K.; Giese, T. J.; Guo, Z.; Li, P.; Lin, C.; McGee, T. D.; Pearlman, D. A.; Radak, B. K.; Tao, Y.; Tsai, H.-C.; Xu, H.; Sherman, W.; York, D. M. Alchemical Binding Free Energy Calculations in AMBER20: Advances and Best Practices for Drug Discovery. J. Chem. Inf. Model. 2020, DOI: 10.1021/acs.jcim.0c00613.
- (27) Chen, H.; Maia, J. D. C.; Radak, B. K.; Hardy, D. J.; Cai, W.; Chipot, C.; Tajkhorshid, E. Boosting Free-Energy Perturbation Calculations with GPU-Accelerated NAMD. *J. Chem. Inf. Model.* **2020**, DOI: 10.1021/acs.jcim.0c00745.
- (28) Zheng, Z.; Borbulevych, O. Y.; Liu, H.; Deng, J.; Martin, R. I.; Westerhoff, L. M. Movable Type Software for Fast Free Energy-Based Virtual Screening: Protocol Development, Deployment, Validation, and Assessment. *J. Chem. Inf. Model.* **2020**, DOI: 10.1021/acs.jcim.0c00618.
- (29) Senapathi, T.; Suruzhon, M.; Barnett, C. B.; Essex, J.; Naidoo, K. J. BRIDGE: An Open Platform for Reproducible High-Throughput Free Energy Simulations. *J. Chem. Inf. Model.* **2020**, DOI: 10.1021/acs.jcim.0c00206.